When Algorithms Discriminate: Tech Bias in Justice Systems

Outcome Document | CSW70 NGO Side Event

11 March 2026 | New York

Convened and moderated by **Caitlin Kraft-Buchman**, CEO/Founder, Women At The Table

Panelists:

- **Laura Nyirinkindi** — Member, UN Working Group on Discrimination Against Women and Girls; Africa Regional Vice-President, International Federation of Women Lawyers (FIDA)
- **Esther Eghobamien-Mshelia** — Member, Committee on the Elimination of Discrimination Against Women (CEDAW)
- **Fernanda K. Martins** — Director of Strategy and Advocacy, Fundación Multitudes

Grounded in the expert paper "Gender Bias in Judicial Algorithms: A Global Analysis of Algorithmic Discrimination," commissioned for the CSW70 Expert Group Meeting on ensuring access to justice for women and girls.

I. THE CORE ARGUMENT: BIAS LAUNDERED, NOT ELIMINATED

The promise of algorithmic justice rests on a seductive premise: that mathematical models can eliminate human bias from judicial decision-making. Courts across at least 20 U.S. states, multiple Canadian provinces, several Australian territories, and pilot programs in the Netherlands, Germany, and France have embraced this vision, deploying automated risk assessment tools for bail determinations, sentencing recommendations, and case management.

The evidence presented at this session — drawn from the expert paper and from the direct testimony of panelists working across UN human rights mechanisms, treaty bodies, and civil society in the Global South — demonstrates that this promise is not just unfulfilled. It is inverted.

Algorithmic systems do not eliminate structural discrimination. They encode it, give it a scientific credential, and scale it — making bias harder to see, harder to challenge, and harder to dismantle.

The algorithm did not rise above the judge's prejudice. It took notes.

As Laura Nyirinkindi framed it from the perspective of the Working Group on Discrimination Against Women and Girls: the Working Group's mandate exists precisely because structural discrimination against women is not incidental — it is built into systems. Algorithmic tools are the newest iteration of that pattern, and they are not exempt from international human rights law.

This is not a technology problem with a technology fix. It is a structural discrimination problem that technology is amplifying at scale.

II. HOW IT WORKS: THREE PATHWAYS OF ALGORITHMIC BIAS

The expert paper documents three primary technical pathways through which gender bias enters judicial algorithms. The panel discussion grounded each of these in lived institutional reality.

Training Data: The Past Encoded as Future

Algorithms learn from historical court decisions — decisions shaped by decades of gender stereotypes. In Brazil, research using BERT language models detects gender bias in court decisions with 88.86% accuracy, revealing systematic patterns where women are characterized through emotional language while men's actions are described through situational factors. When these linguistic patterns become training data, the systems learn to associate women with unpredictability and men with circumstantial violence.

Fernanda Martins situated this within a broader "justice chain": AI enters the system before a judge writes a decision — through surveillance systems, predictive policing tools, and risk analysis used by security institutions. By the time a case reaches a courtroom, technological systems may already have shaped who was identified as suspicious, who was investigated, and who was prosecuted.

All a predictive algorithm is doing is taking the past and projecting it into the future. It is not predicting — it is reproducing.

Proxy Variables: Discrimination Without Naming It

Even when gender is removed as an explicit variable, algorithms discriminate through proxies. Employment status, residential stability, relationship history, and educational background all correlate strongly with gendered social roles. A woman with interrupted employment due to caregiving is scored as "less stable." A woman in an under-resourced neighbourhood is scored as "higher risk." The algorithm never mentions gender — and yet gender is everywhere in it.

Laura Nyirinkindi was direct on this point: while the system may never explicitly reference gender, gender is embedded in the data, and therefore in the outcomes. This is a classic example of indirect discrimination — where seemingly neutral criteria produce disproportionate harm for women.

European implementations reveal culturally specific manifestations: employment gaps penalized in German systems disproportionately affect women with caregiving responsibilities; "social support" variables in Dutch systems may score women's family networks as dependency while treating men's peer networks as positive social capital; financial stability measures across EU systems reflect wage gaps.

The Credibility Gap — Now Automated

Women already face significantly lower credibility ratings from both male and female judges, particularly in sexual assault cases. When algorithms train on historical decisions embedding these credibility biases, they systematize and scale that gap. Natural language processing systems trained on court transcripts learn to associate women's speech patterns with uncertainty markers. Sentiment analysis tools may interpret women's emotional expressions as less reliable.

Esther Eghobamien-Mshelia illustrated this with a striking example from South Africa: police records did not have rape explicitly listed as a crime category — reported rapes were coded as "assault" or "physical violence." When these records become algorithmic training data, the system inherits not just bias but entire categories of erasure. The definitional frameworks that shape how violence against women is recorded were constructed without women's input — and those frameworks now power the algorithms.

III. THE IMPOSSIBILITY THEOREM: WHY FAIRNESS IS A POLITICAL CHOICE

Computer scientists have proven mathematically that no algorithm can simultaneously achieve predictive parity, equal false-positive rates, and equal false-negative rates when base rates differ between groups. This is not a technical limitation to be engineered away. It is a mathematical proof — the impossibility theorem.

There are 21 different mathematical definitions of fairness, and they cannot all be satisfied at the same time. Choices must be made: fair to whom, measured how, at whose expense. These choices are inherently political, not technical — and they cannot be delegated to developers or data scientists.

This has a direct consequence for the governance conversation. When accuracy is treated as the sole measure of system quality — as it remains in most data science — there is no room in the equation for fairness or transparency. The panel emphasized that a system designed to accept slightly less accuracy in exchange for greater fairness and transparency is not a concession. It is a design constraint that can be built in — but only if it is demanded.

IV. WHO BEARS THE COST: INTERSECTIONAL EVIDENCE

The discriminatory outcomes documented in the expert paper and discussed during the session are not evenly distributed. They compound along intersectional lines.

Documented patterns of algorithmic discrimination include:

- **United States:** Women rated "high risk" by the COMPAS algorithm had less than half the actual reoffending rate of men rated "high risk" (25% versus 52%) — systematic overprediction of women's recidivism. Black defendants were 77% more likely to be rated high risk for violent recidivism even after controlling for age and gender.

- **Netherlands:** The HART system rated Moroccan and Turkish immigrant women 40% higher risk than Dutch women with identical criminal histories.

- **Germany:** Pilot data from Düsseldorf courts showed women receiving 23% higher risk scores than men for identical property crimes. Trans women were systematically classified using male risk factors regardless of gender identity.

- **Australia:** Aboriginal women in New South Wales were systematically rated higher risk despite lower actual reoffending rates — racial and gender bias compounding through algorithmic amplification.

- **Canada:** Ontario's SIR-R1 system rated Indigenous women 35% higher risk despite significantly lower violent reoffending rates.

- **United Kingdom:** Transgender defendants faced 60% higher risk ratings than cisgender defendants with identical criminal histories.

Esther Eghobamien-Mshelia connected these patterns to a deeper structural analysis: incarceration itself has not been a solution to gender-based violence. When a perpetrator is incarcerated, the woman often becomes the sole provider, carrying the economic burden alongside the stigma. The algorithmic systems that funnel people toward incarceration rarely account for these downstream effects on women and families.

Fernanda Martins underscored that in Latin America, where Brazil has one of the largest prison populations in the world and nearly 70% of incarcerated people are Black, automation is entering institutions that already produce unequal outcomes. There is a real risk that technological tools reproduce and create patterns of discrimination while presenting them as objective or efficient decisions.

V. WHAT INTERNATIONAL LAW REQUIRES — AND WHAT IS MISSING

The CEDAW Framework

The Convention on the Elimination of All Forms of Discrimination Against Women, with 189 States Parties, provides binding obligations on access to justice. General Recommendation No. 33 on women's access to justice speaks directly to States'

obligations when their legal systems produce discriminatory outcomes — algorithmic or otherwise.

Esther Eghobamien-Mshelia was explicit: the Convention's 16 operative articles and its General Recommendations provide more than adequate standards to respond to the challenges posed by algorithmic discrimination. What is lacking is political will — both from governments and from technology companies. The frameworks exist. The resourcing, expertise, enforcement, and gender-specific attention do not.

A critical development raised during the session: CEDAW is currently developing **GR41 on stereotyping**, which will directly address stereotyping in the digital era — including how algorithmic systems reinvent and scale gender stereotypes. The draft is open for consultation.

The Working Group on Discrimination Against Women and Girls

The WG DAW will present its thematic report on gender equality in digital spaces and artificial intelligence to the Human Rights Council in June 2026. Laura Nyirinkindi shared that consultations across regions have revealed wide variation — some countries engaging as consumers of AI, others as producers, others as highly knowledgeable, and others with minimal general awareness. But across every region, AI-related harms compound along intersectional lines.

A key finding from the WG DAW's work: while algorithmic tools are not exempt from international human rights law, they are increasingly exempt from national human rights frameworks. This gap between the international normative standard and national implementation is where discrimination takes root.

Regulatory Landscape

The EU AI Act represents the most comprehensive regulatory framework, classifying criminal justice applications as high-risk and requiring mandatory bias assessment with penalties reaching €35 million or 7% of global turnover. Germany's Federal Constitutional Court has held that algorithmic decision-making must respect human dignity and cannot reduce individuals to data points.

However, key implementation gaps persist across all jurisdictions: most frameworks remain voluntary, gender-specific provisions receive less attention than racial bias, oversight mechanisms lack enforcement power, and international coordination remains minimal. In Latin America, as Fernanda Martins noted, AI regulation discussions are focused on risk frameworks without human rights, gender, or intersectional lenses — and the lobbying power of major platforms is actively blocking progress.

VI. ACCOUNTABILITY LEVERS AND CALLS TO ACTION

The session moved deliberately from analysis to action. The panel identified concrete accountability mechanisms — several of which emerged from the discussion itself rather than the prepared materials.

Licensing and Relicensing as Leverage

Esther Eghobamien-Mshelia identified a mechanism that has been underexplored: the licensing and relicensing of technology platforms and service providers by national regulators. When companies must renew their operating licenses, governments have a direct point of leverage to demand compliance with human rights obligations — including algorithmic accountability. This mechanism already exists within regulatory frameworks; it has simply not been deployed for algorithmic discrimination. The Working Group on Business and Human Rights is willing to collaborate on this approach.

Cascading CEDAW to Cities and Communities

CEDAW's global campaign to cascade Convention obligations to the city and community level through ordinances, bylaws, and local edicts offers a concrete pathway for making algorithmic accountability justiciable at the local level. When CEDAW obligations are localized, women can seek protection and remediation within their own communities rather than navigating prohibitively expensive national legal

systems. Technology violations — including algorithmic discrimination — can be monitored, tracked, and held accountable within the mandate of local authorities.

The WG DAW Communications Mechanism

The Working Group on Discrimination Against Women and Girls can receive credible allegations of violations and write communications to governments requesting information on measures taken. This mechanism is available to individuals and organizations and is accessible through the Working Group's website.

Transparency and the Right to Know

The session converged on a principle that cuts across all jurisdictions: a right to know when an automated system has contributed to a decision, how it reached that decision, and on what basis it can be challenged. This encompasses transparency in algorithmic design, mandatory disclosure when algorithmic tools influence judicial decisions, and meaningful avenues for affected individuals to contest those decisions.

Capacity Building Across the Justice Chain

Fernanda Martins emphasized the necessity of building capacity across the entire justice system — from police to lawyers to judges to platform operators — to understand how algorithmic tools function and where they introduce bias. This includes gender-responsive training for developers and algorithmic literacy for judicial actors. Technological deployment is advancing faster than institutional capacity to govern these systems in line with human rights obligations.

Design from Inception

The panel was unanimous: gender-responsive design cannot be retrofitted. It must be embedded from ideation through deployment. This means women and affected communities involved in problem definition, algorithm design, bias testing, and oversight — not consulted after key choices are made. As Esther Eghobamien-Mshelia put it: without parity in leadership, women cannot challenge

the codes, the models, or the services. When there are too few women to change the system, even those present cannot shift the formula.

VII. EMERGING FRAMEWORKS AND NEXT STEPS

Immediate Opportunities

- **CEDAW GR41 on Stereotyping** — currently open for consultation. Civil society organizations should submit inputs addressing algorithmic stereotyping specifically.

- **WG DAW Thematic Report on AI and Gender Equality** — to be presented to the Human Rights Council in June 2026. This report will serve as a critical advocacy tool. NGOs should prepare to use it in national advocacy and shadow reporting.

- **CEDAW Periodic Review** — civil society advocates should ensure that algorithmic discrimination appears in shadow reports and concluding observations. The Committee has signalled its readiness to engage on this issue.

- **National AI Regulation** — wherever AI governance legislation is under development, advocates should push for human rights-based and gender-responsive frameworks, not risk-only approaches. The gap between international standards and national implementation is where the most urgent work lies.

Structural Priorities

- Mandatory algorithmic audits for all government AI systems, with gender-disaggregated results and public disclosure
- Independent external monitoring with meaningful access to systems and data

- Community oversight mechanisms that include affected populations in algorithm design and evaluation
- International coordination on judicial algorithm standards — currently minimal despite cross-border implications
- Gender-responsive AI procurement frameworks for municipalities and local governments

---

VIII. THE FUNDAMENTAL QUESTION

The question confronting judicial systems worldwide is not whether courts will use algorithms. They already are. The question is whether they will implement them responsibly — with adequate protections for gender equality and human dignity — or whether the promise of efficiency will be allowed to override the demands of justice.

The evidence presented at this session demonstrates that algorithmic neutrality is a myth. Bias is not a bug to be fixed but a structural feature requiring deliberate political and ethical choices. Those choices — about what fairness means, whose data counts, and who bears the cost of error — cannot be delegated to technical experts. They must involve affected communities, legal scholars, human rights mechanisms, and democratic oversight processes.

The international human rights framework is not silent on this. CEDAW, the WG DAW mandate, the EU AI Act, and constitutional protections across jurisdictions all provide grounds for demanding accountability. What is missing is not the law. It is the political will to enforce it — and the insistence that speed and efficiency cannot come at the cost of equality and justice.

Operational, not aspirational. Designed with, not designed for.

About the Expert Paper "Gender Bias in Judicial Algorithms: A Global Analysis of Algorithmic Discrimination" was commissioned for the CSW70 Expert Group Meeting and prepared by Caitlin Kraft-Buchman, CEO/Founder, Women At The Table.

About the Convening Organization Women At The Table works on systems change in technology governance — ensuring women have genuine authority in the decisions that shape how technology gets built, deployed, and held accountable.